

Ref: CLME'2011\_0705A

## A INSOLAÇÃO COMO PARÂMETRO DE ENTRADA EM MODELO BASEADO EM REDES NEURONAIIS PARA PREVISÃO DE SÉRIE TEMPORAL DO TURISMO

João Paulo Teixeira<sup>\*1</sup>, Paula Odete Fernandes<sup>2</sup>

<sup>1</sup>Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança - Bragança, Portugal

<sup>2</sup>Núcleo de Estudos em Ciências Empresariais, UBI - Covilhã, Portugal

\*Email: joaopt@ipb.pt

### RESUMO

*A série temporal de Dormidas nos estabelecimentos hoteleiros é usada como variável indicadora da actividade de turismo. Esta série tem vindo a ser objecto de modelação para efeitos de previsão pelos autores, usando modelos ARIMA e modelos baseados em redes neuronais artificiais.*

*Em trabalhos anteriores o modelo usado tem sido baseado em RNA do tipo feed-forward, com uma camada escondida para fazer a previsão da série para o futuro. Este modelo tem na saída a variável a prever (número de dormidas para o próximo mês) e na entrada tem sido usado o valor das dormidas dos 12 meses anteriores. Numa evolução mais recente do modelo foram usadas na entrada apenas duas variáveis indicadoras do ano e do mês para o qual se pretende fazer a previsão.*

*Neste trabalho apresenta-se o estudo e os resultados da modelação da série temporal de Dormidas na região Norte de Portugal combinando os modelos anteriores com uma nova variável que regista o número de horas de sol por mês. Com a introdução desta nova variável pretende-se inserir no modelo a apetência causada pelo estado do tempo em turistas com saídas não marcadas previamente.*

*Da aplicação do modelo pode-se referir que efectivamente, nos modelos comparados, o erro relativo médio é reduzido, cerca de 0.5 % com a introdução desta nova variável.*

**Palavras-Chave:** Modelos de Previsão, Redes Neuronalis, Séries Temporais, Turismo.

### 1. INTRODUÇÃO

Reconhecido como um dos sectores estratégicos da economia portuguesa, o turismo tem vindo a desempenhar um papel fundamental para o desenvolvimento do país. Desde os últimos anos que se tem assistido a um crescimento turístico continuado, embora com algumas inflexões pontuais, que podem ser consequência de alguma instabilidade económica. O ano recente de 2009 ficou assinalado pela crise financeira mundial e pela ocorrência da pandemia da gripe H1N1. À semelhança de crises anteriores, também se verificou que o turismo, em Portugal, foi suportado pelo turismo interno.

Contudo e tendo como base os dados da Conta Satélite do Turismo [Turismo de Portugal, 2010a], em 2009 para Portugal, na óptica da oferta, o Valor Acrescentado Gerado pelo

Turismo (VAGT) contribuiu em 4,8% para o Valor Acrescentado Bruto (VAB) da economia, representando uma das principais actividades económicas do país. Já na óptica da Procura Turística, o peso do Consumo Turístico Interior na economia nacional representou 10,0% do PIB. O saldo da Balança Turística contribuiu positivamente para a diminuição do défice da Balança Corrente, registando, em 2009, um valor positivo de 4,2 mil milhões €. Em 2009, as Receitas Turísticas apresentam valores superiores comparativamente aos principais sectores exportadores, concentrando uma quota global de 14,6% no Total das Exportações de Bens e Serviços.

À semelhança de Portugal também o sector turístico na Região Norte de Portugal tem crescido de uma forma continuada. Em 2009, os proveitos auferidos situaram-se nos 207,6 milhões de euros e a região recebeu 2,5 milhões de hóspedes que originaram 4,4 milhões de dormidas, sendo que a sua permanência média foi de 1,7 noites por hóspede. O mercado interno foi responsável por 59% das dormidas registadas na região [Turismo de Portugal, 2010b].

Neste contexto, pelos números apresentados anteriormente, torna-se pertinente as diferentes entidades envolvidas no sector turístico planearem cuidadosamente as acções a tomar para atrair e fidelizar os turistas que se deslocam à região. Que as decisões sejam suportadas num sistema de actividades diferenciadas que se reforcem mutuamente e que sejam ajustadas face aos ambientes contextuais futuros. Assim, torna-se vital efectuar previsões da procura turística com alguma precisão e encontrar as diferentes variáveis que permitem explicar o comportamento da procura turística.

Nesta linha, os autores têm trabalhado com diferentes modelos de previsão com distintas características, distintos níveis de complexidade e têm utilizado variáveis económicas explicativas da procura turística [Fernandes *et al.*, 2008; Fernandes *et al.*, 2009b]. A série objecto de estudo e que tem sido modelada é a série temporal de turismo: “Dormidas Mensais nos Estabelecimentos Hoteleiros na Região Norte de Portugal”. Esta série apresenta um comportamento sazonal, marcada por picos mais elevados nos meses de Verão e mais baixos nos meses de Inverno. Dado este comportamento e porque o turismo em Portugal sempre esteve muito associado ao produto turístico balnear, procurou-se com o presente trabalho avaliar a influência do número de horas de sol aberto na variação sazonal, acreditando que os fluxos dentro do país ou até mesmo de Espanha para Portugal possam ser influenciados pelo estado do tempo, aqui associado à variável Insolação. Para tal utilizou-se o modelo tendo por base as Redes Neurais Artificiais e as séries Dormidas Mensais nos Estabelecimentos Hoteleiros e o Número de Horas Mensais de Insolação. Os dados observados, para as séries referidas consideram o período compreendido entre Janeiro de 1987 e Dezembro de 2009, correspondendo assim a 276 observações mensais, ao longo de 23 anos.

O presente trabalho encontra-se estruturado da seguinte forma: depois de uma breve introdução, procede-se à apresentação e análise da série temporal “Dormidas Mensais nos Estabelecimentos Hoteleiros na Região Norte de Portugal” e da variável Insolação incluída no modelo. Segue-se uma descrição dos modelos aqui comparados e posteriormente, apresenta-se a previsão da procura turística, com base nos modelos construídos, para os anos 2008 e 2009, analisando-se a *performance* dos mesmos. Por último, serão apresentadas as principais conclusões do estudo.

## 2. APRESENTAÇÃO E ANÁLISE DO COMPORTAMENTO DAS SÉRIES TEMPORAIS

A série em análise refere-se à região Norte de Portugal e apresenta-se na Figura 1 a sua evolução mensal desde Janeiro de 1987 até Dezembro 2009 (INE, 1987-2010). Como se pode observar a série tem um comportamento sazonal com maior procura nos meses quentes e menos nos meses de inverno. Sabemos também que o calor está muito ligado às horas de sol, que será experimentado neste trabalho. Pode ainda observar-se que esta teve um ligeiro crescimento no final da década de 80, uma estagnação no início da década de 90, arrastada pela crise económica desse período, e voltou a crescer em finais da década de 90 e até 2002. Depois esteve 4 anos a diminuir voltando a crescer significativamente a partir de 2005. Este crescimento pode ser resultado de investimentos realizados em variáveis de marketing e em eventos mundiais e europeus que permitiram promover a região não só a nível nacional mas também a uma dimensão internacional.

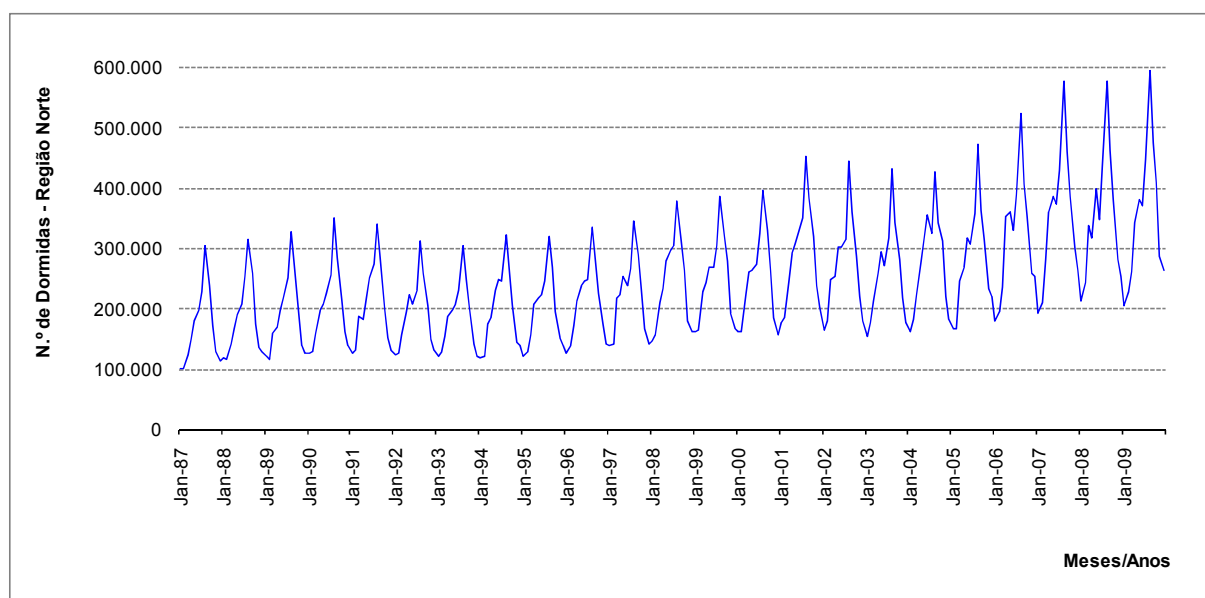


Figura 1: Série temporal do número de Dormidas na região Norte de Portugal.

No presente trabalho, procura-se explorar a utilidade de uma outra variável com características climáticas: Número de Horas Mensais de Sol, denominada de Insolação, no sentido de averiguar se a mesma se encontra relacionada com a procura turística na região Norte de Portugal. Pois como é sabido, Portugal tem sido promovido como sendo um país que oferece o produto turístico sol e mar, o que atrai milhares de turistas.

Os valores para a variável Insolação foram registados na Estação de Meteorologia de Porto/Serra do Pilar entre Janeiro de 1987 e Dezembro de 2005, e pela estação meteorológica de Porto/Pedras Rubras entre Janeiro de 2006 e Dezembro de 2009. Os valores de Insolação não estão disponíveis em nenhuma das estações para o período completo em análise. Estes dados de Insolação foram facultados pelo Instituto de Meteorologia de Portugal.

Pela análise da variável Insolação apresentada na Figura 2, pode observar-se que esta também apresenta um comportamento sazonal, registando valores elevados na época alta (entre Abril e Setembro) e valores mais baixos na época baixa (Outubro a Março). Na Figura 3 sobrepõe-se

a representação da série de Dormidas e da variável Insolação (por uma questão de escala a Insolação representada nesta figura vem multiplicada por 1000). A pertinência da utilização desta variável parece evidente pela sobreposição da sua sazonalidade com a da série em estudo Dormidas. O coeficiente de correlação entre a série Dormidas e a variável Insolação é de 0.571. Este valor parece também confirmar a importância desta variável.

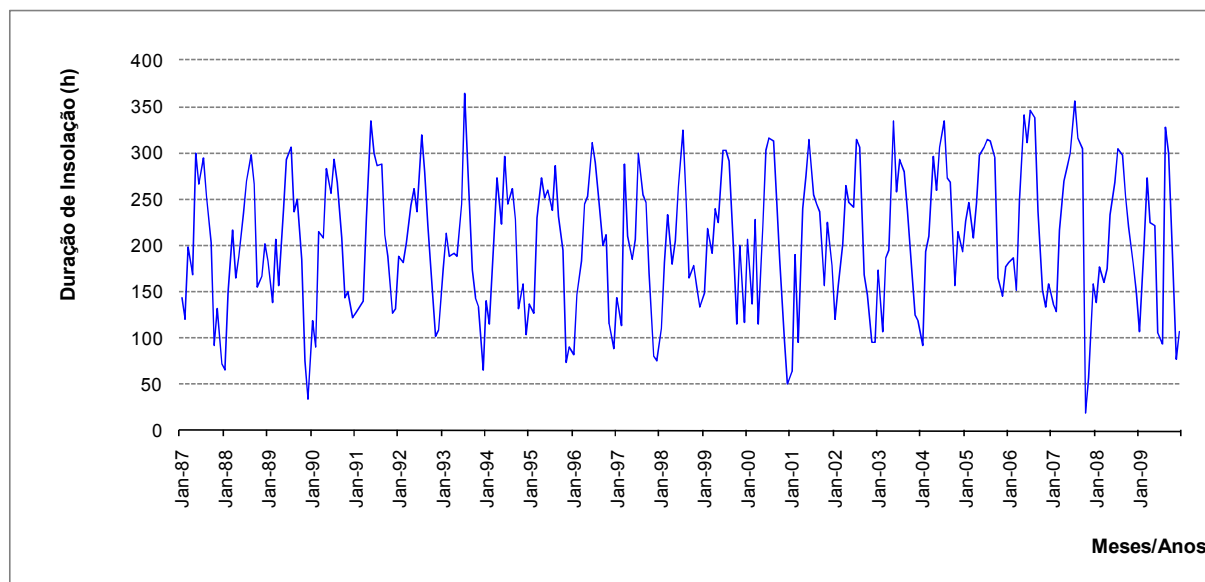


Figura 2: Insolação.

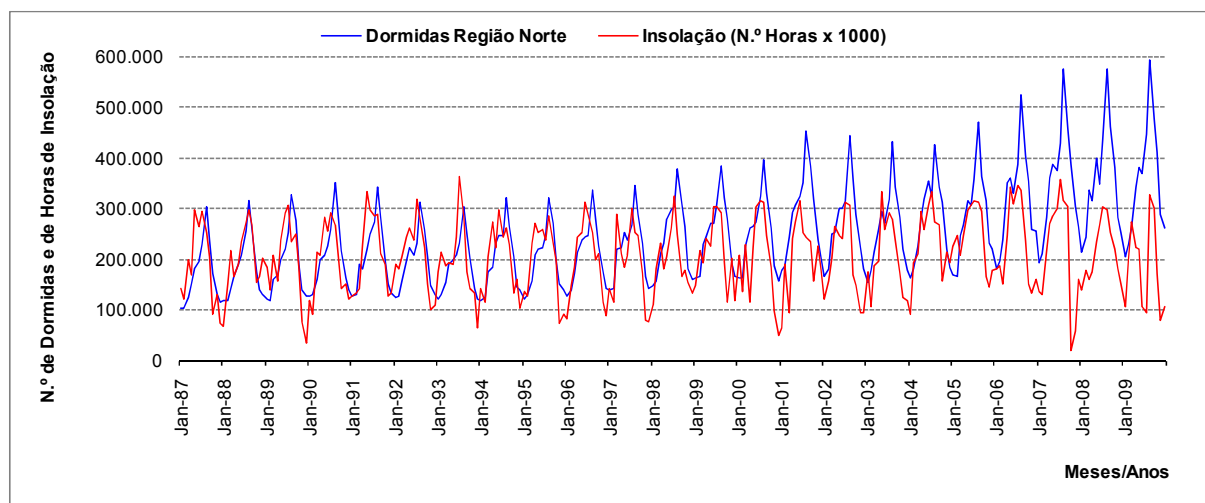


Figura 3: Série temporal Dormidas e o número de horas mensais de sol (Insolação).

Os valores de Insolação mensal foram registados nas Estações Meteorológicas do Porto na Serra do Pilar e Pedras Rubras. Optou-se por esta localização das estações meteorológicas por duas razões: a primeira prende-se com o facto do destino Grande Porto deter um peso determinante na procura turística, por si só representa mais de metade das dormidas registadas, ao nível de NUT III da Região Norte; e por outro lado, pelo facto das horas de Insolação anuais variarem entre as 2.200h e as 2.600h distribuídas pela região Norte,

registando-se uma uniformização nos diferentes locais da região Norte, tal como se pode observar na Figura 4. Assim, pelo exposto considera-se que o registo da Insolação neste local é representativo da variável do estado do tempo para a região Norte quando relacionada com o número de Dormidas.

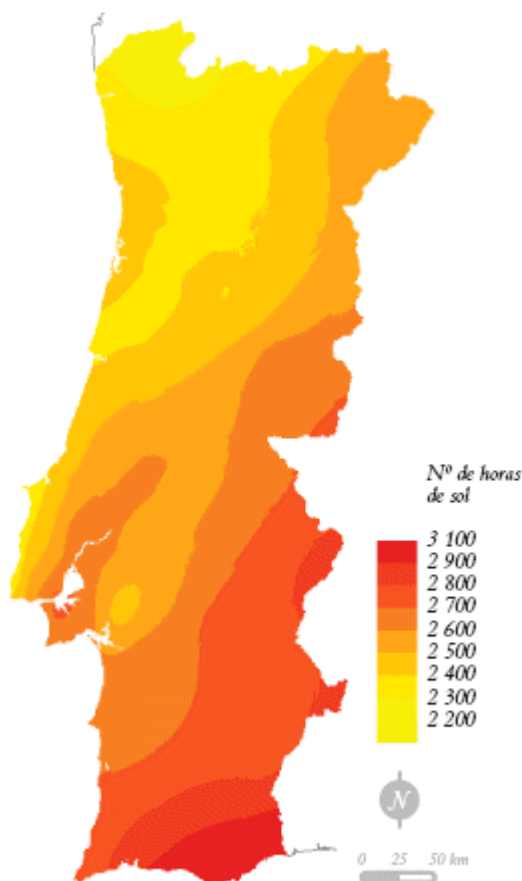


Figura 4: Distribuição do número de horas de Insolação, anuais em Portugal Continental [<http://www.igeo.pt>].

A utilidade desta variável de Insolação na ajuda à previsão do número de dormidas de uma região para o mês seguinte é importante, para já em termos conceptuais, apenas, uma vez que esta variável também não é conhecida previamente para efeitos de previsão para o próximo mês, numa situação real.

### 3. MODELOS DE REDES NEURONAIAS

Foram usados 4 modelos baseados em Redes Neurais Artificiais (RNA), dois deles sem a utilização da variável Insolação e os outros dois semelhantes aos primeiros mas com a Insolação na entrada. O objectivo consiste em verificar se melhora o desempenho de previsão dos modelos com a Insolação.

Em todos os modelos pretende-se fazer a previsão do número de dormidas para o mês seguinte. Assim a saída da rede neuronal consiste apenas num nó. A entrada é diferente em cada modelo.

A série Dormidas foi dividida em três conjuntos. Um primeiro que é usado no treino da rede neuronal, denominado de conjunto de Treino, contendo os valores desde Janeiro de 1987 (ou de 1988, quando o modelo necessita dos valores dos últimos 12 meses na entrada, como é o caso dos modelos A e C, a seguir apresentados) até Dezembro de 2007. Isto representa um total de 252 ou 240 (para os modelos A e C) pares de entrada saída disponíveis para treino. Um segundo conjunto de Validação usado para interromper o treino por um processo de validação cruzada. Este conjunto contém os meses de Janeiro a Dezembro de 2008. O terceiro conjunto consiste no conjunto de Teste e é usado apenas para avaliar o desempenho do modelo no final, não sendo usado no treino. Este conjunto é composto pelos meses de Janeiro a Dezembro de 2009.

### **Descrição do Modelo A**

Este modelo tem vindo a ser usado pelos autores com sucesso para a previsão da série Dormidas de diversas regiões [Fernandes *et al.*, 2008; Teixeira e Fernandes, 2008]. Por isso tem vindo também a ser usado pelos autores para comparação com outros modelos estudados [Fernandes *et al.*, 2009a; Fernandes *et al.*, 2009b; Teixeira e Fernandes, 2010], pelo que também se têm referido a este modelo como o modelo clássico.

Este modelo consiste numa rede neuronal do tipo *feed-forward* com 12 nós na entrada que recebem as valores da série dos últimos 12 meses, uma camada escondida com 6 nós, e um nó na saída com o valor previsto para o mês seguinte. A função de activação na camada escondida é a tangente hiperbólica e na saída a função linear.

### **Descrição do Modelo B**

Este modelo foi aplicado com sucesso às séries da região Norte e Centro [Fernandes e Teixeira, 2009] e posteriormente à série da região Madeira [Teixeira e Fernandes, 2010]. Consiste em usar apenas variáveis que permitem ao modelo ter informação sobre a evolução temporal e desta forma inferir a tendência.

Concretamente o modelo consiste numa rede neuronal *feed-forward* com dois nós na entrada para codificar o ano e o mês. O ano é sequencial a começa em 1 para o primeiro ano. O mês também é sequencial entre 1 e 12 para os meses de Janeiro a Dezembro. A camada escondida contém 12 nós. Este modelo terá que aprender nas suas variáveis internas (pesos e desvios) a evolução sazonal da série, pelo que necessita de mais nós na camada escondida. A camada de saída contém apenas um nó para a previsão do valor de dormidas para o mês em causa. Na camada escondida a função de activação é a tangente hiperbólica e na camada de saída é uma função linear.

### **Descrição do Modelo C**

Este modelo é igual ao modelo A, com mais um nó na camada de entrada para incluir a variável Insolação do mês a prever.

Da comparação dos resultados deste modelo com os do modelo A espera-se avaliar a pertinência da utilização da variável Insolação.

## Descrição do Modelo D

O modelo D é semelhante ao modelo B contendo mais um nó na entrada para a variável Insolação. Neste modelo o número de nós da camada escondida foi otimizado experimentalmente para 20 nós.

Também a comparação dos resultados destes dois modelos, B e D, permitirá avaliar a utilidade desta nova variável.

## 4. DESEMPENHO DOS MODELOS

Para avaliar o desempenho dos modelos foi medido o erro relativo médio, dado pela eq. 1, para o conjunto de Validação e para o conjunto de Teste, separadamente.

$$ER(\%) = \frac{1}{N} \sum_{i=1}^N \frac{|T_i - S_i|}{T_i} (\times 100) \quad [\text{Eq. 1}]$$

Em que N é a dimensão do conjunto,  $T_i$  o valor alvo e  $S_i$  o valor da saída.

Apesar de se apresentarem os erros relativos para os dois conjuntos de Validação e Teste, será mais relevante o desempenho do modelo no conjunto de Teste do que no conjunto de Validação, uma vez que não foi usado durante o processo de treino da rede neuronal.

Seguidamente apresenta-se o desempenho de cada modelo por comparação da serie alvo (Dormidas) com os valores previstos.

A Figura 5 mostra o desempenho do modelo A: Nesta figura os valores previstos estão a vermelho e os dois últimos ciclos correspondem aos meses dos anos de 2008 e 2009, que coincidem com os conjuntos de Validação e Teste, respectivamente.

As Figuras 6, 7 e 8, apresentam o desempenho dos modelos B, C e D, respectivamente.

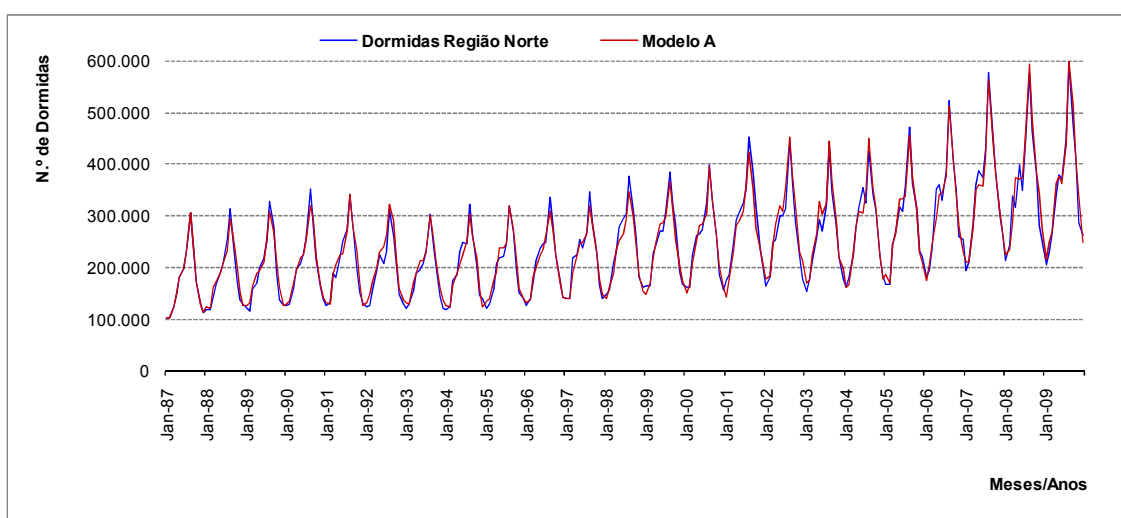


Figura 5: Valores reais das Dormidas a azul e valores previstos pelo modelo A, a vermelho.

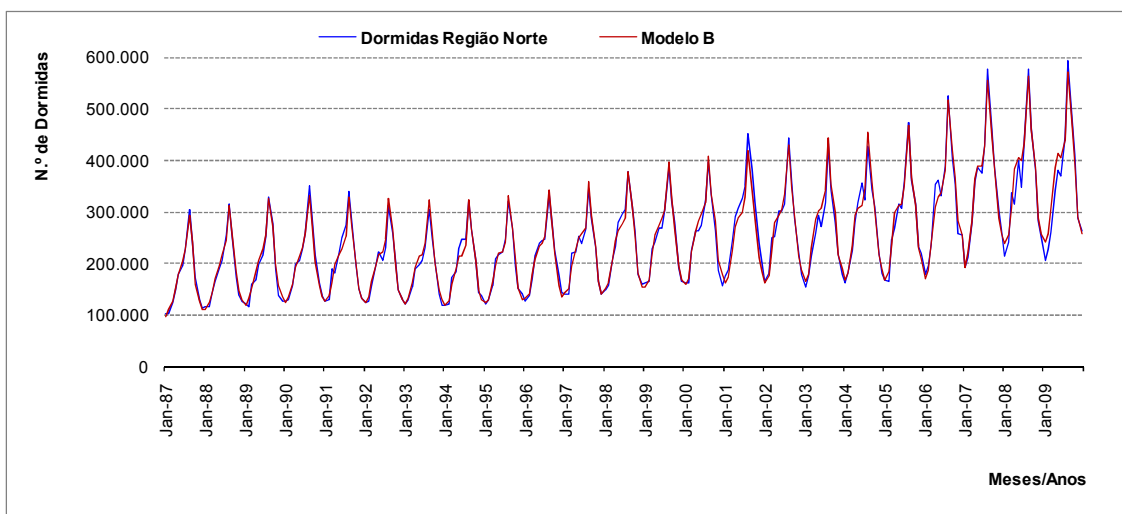


Figura 6: Valores reais das Dormidas a azul e valores previstos pelo modelo B, a vermelho.

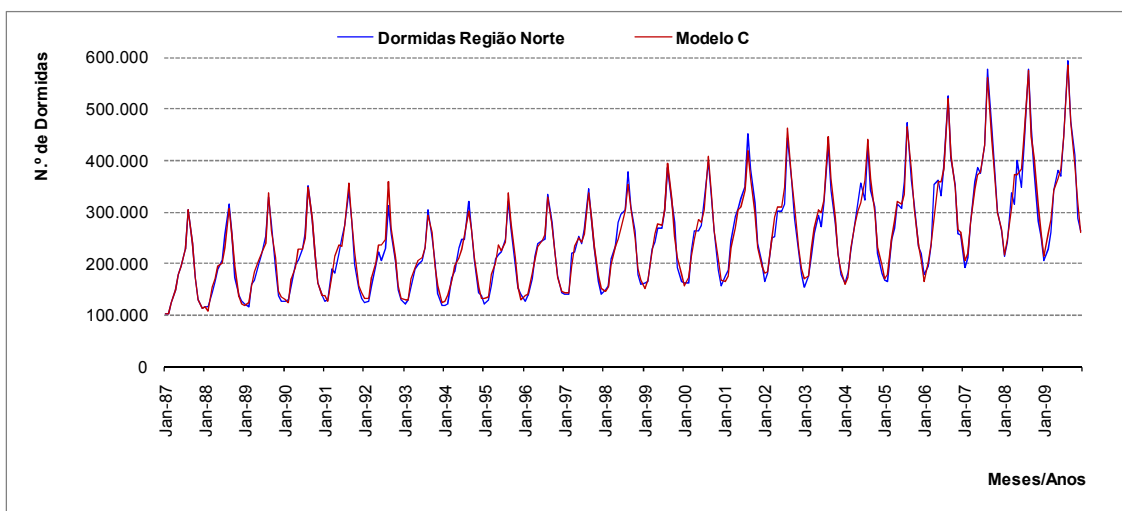


Figura 7: Valores reais das Dormidas a azul e valores previstos pelo modelo C, a vermelho.

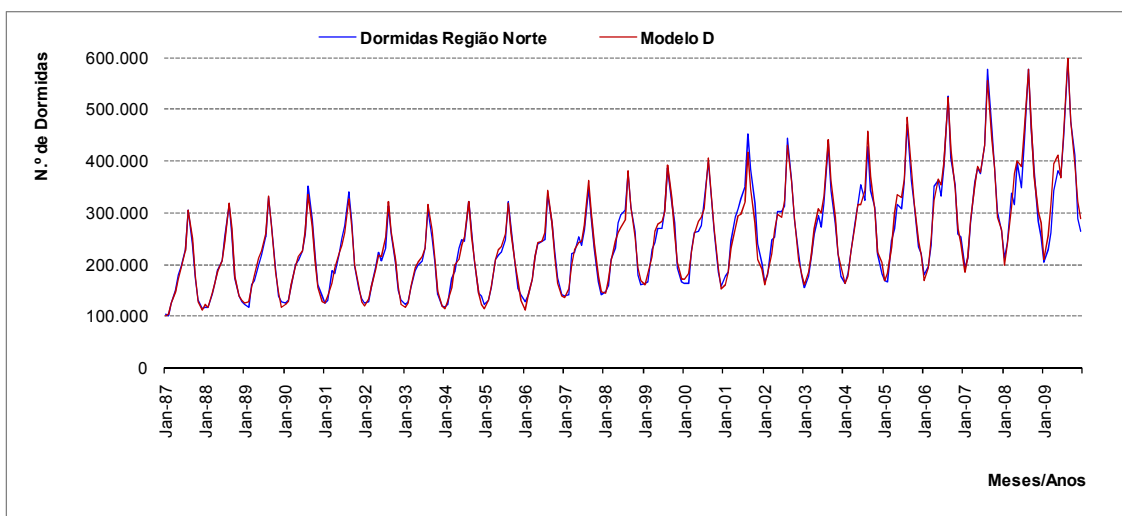


Figura 8: Valores reais das Dormidas a azul e valores previstos pelo modelo D, a vermelho.



Comparando o desempenho dos 4 modelos, onde se observa que todos seguem o comportamento da série com os valores reais, apresenta-se na Figura 9 a previsão realizada pelos modelos e o valor alvo da série Dormidas, no período de 2008 a 2009, correspondentes aos conjuntos de Validação e Teste, respectivamente. Pela visualização da figura verifica-se para o período em estudo, qualquer um dos modelos segue o comportamento apresentado pela série original, observando-se ainda, graficamente, que os modelos A e C, são os que seguem de mais perto a série original.

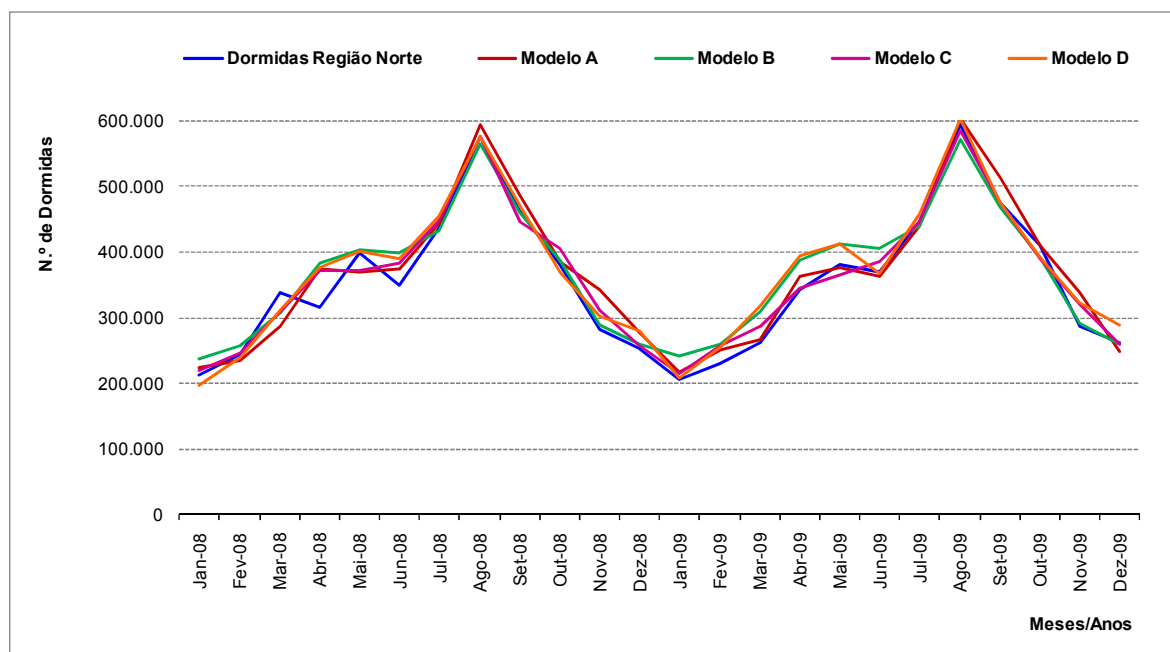


Figura 9: Valores reais das Dormidas e valores previstos pelos modelos A, B, C e D.

A medição do erro relativo nos conjuntos de Validação e Teste para os 4 modelos é apresentada na Tabela 1. Tendo por base, para a análise, o conjunto de teste, como já se referiu este serve para medir a *performance* de cada modelo, uma vez que os valores nunca foram usados aquando do treino, pode observar-se que os modelos que apresentaram melhor *performance* foram os modelos A e C. Já anteriormente, na análise gráfica, se tinha chegado à mesma conclusão. De sublinhar, uma vez que era objectivo deste estudo, que quando se introduziu a nova variável - Insolação - os modelos produziram valores mais próximos dos registados efectivamente, tal como se pode observar pela diminuição do Erro Relativo (ER) alcançado para os modelos C e D. No entanto e segundo a classificação proposta por Lewis (1982), para o indicador Erro Relativo, como critério de análise da qualidade dos valores previstos, pode dizer-se que os modelos conseguiram produzir previsões de precisão elevada, uma vez que o ER apresenta valores inferiores a 10%.

Tabela 1. Erro Relativo.

Conjuntos/Erro (%)	Modelo A	Modelo B	Modelo C	Modelo D
<b>Validação</b>	8,19	6,12	5,98	6,24
<b>Teste</b>	4,93	7,78	4,45	7,30

## 5. CONCLUSÕES

Com o presente trabalho procurou-se desenvolver um modelo que permita realizar a previsão, com precisão elevada, do número de dormidas mensal de uma região, neste caso da região Norte de Portugal. Este modelo já tem vindo a ser objecto de estudo pelos autores, e neste caso pretendia-se analisar a pertinência para o modelo da inclusão de uma nova variável que consistiu na Insolação mensal (número de horas de sol por mês). Esta variável apresenta uma significativa correlação ( $r=0,571$ ) com a série a prever, e mostra ter a mesma natureza sazonal.

Foram comparados dois modelos desenvolvidos anteriormente pelos autores, com os mesmos modelos incluindo agora esta nova variável na entrada. Da análise dos gráficos apresentados na secção anterior pode concluir-se que os modelos apresentaram valores satisfatórios, conseguindo fazer uma boa previsão da série a avaliar pelo seguimento muito próximo dos valores efectivamente registados nos estabelecimentos hoteleiros. Já olhando para os valores do erro relativo das previsões efectuadas pelos modelos C e D (que incluem a nova variável), no conjunto de Teste, pode concluir-se que esta nova variável veio permitir uma redução do erro no modelo A (modelo clássico) de 4,93% para 4,45%, e uma redução do erro no modelo B, de 7,78% para 7,30%. Em ambos os casos o ganho conseguido com a inclusão desta variável é da ordem de 0,5%.

Finalmente, deve referir-se que os modelos B e D apresentam um erro relativo maior, mas também são modelos com menos variáveis na entrada e portanto mais simples.

## REFERÊNCIAS BIBLIOGRÁFICAS

Fernandes, Paula and Teixeira, João - New Approach of the ANN Methodology for Forecasting Time Series: Use of Time Index. International Conference of Tourism Development and Management. Kos, Greece, 2009.

Fernandes, Paula O.; Monte, Ana Paula e Teixeira, João Paulo - Previsão da procura turística utilizando um modelo não linear. XIII - Congreso Internacional de Investigación en Ciencias Administrativas. Mexico, 2009a.

Fernandes, Paula O.; Teixeira, João Paulo; Ferreira, João José e Azevedo, Susana Garrido - Modelling tourism demand: a comparative study between artificial neural networks and the Box-Jenkins methodology. Romanian Journal of Economic Forecasting. ISSN 1582-6163. 9:3, 2008, p. 30-50.

Fernandes, Paula O.; Teixeira, João Paulo; Ferreira, João José e Azevedo, Susana Garrido - O impacto da variável Páscoa na previsão da procura turística, ASEPELT'09. Covilhã, 2009b.

INE - Anuários Estatísticos do Turismo de 1987 a 2008. Lisboa, 1987-2010.

Instituto Geográfico Português, 2010, <http://www.igeo.pt> (acedido em 16 de Abril de 2011).

Lewis, C.D. - Industrial and Business Forecasting Method, Butterworth Scientific. London, 1982.

Teixeira, João Paulo e Fernandes, Paula O. - Nova abordagem da metodologia de redes neuronais artificiais para a previsão de séries temporais de turismo: a data com índice. Aplicação à Região da Madeira. 16º Congresso da APDR, Funchal, 2010.

Teixeira, João Paulo e Fernandes, Paula O. - Previsão da série temporal turismo com redes neuronais artificiais, Congresso Luso-Moçambicano de Engenharia - CLME. 5.º Maputo, 2008.

Turismo de Portugal, O Turismo em 2009, 2010b, <http://www.turismodeportugal.pt> (acedido em 16 de Abril de 2011).

Turismo de Portugal, O Turismo na Economia: Indicadores de Actividade Turística e Económica em Portugal 2000-2009, 2010a, <http://www.turismodeportugal.pt> (acedido em 16 de Abril de 2011).